Markov Chains

Andreas Klappenecker

Texas A&M University

© 2018 by Andreas Klappenecker. All rights reserved.

A stochastic process $X = \{X(t) : t \in T\}$ is a collection of random variables. The index t usually represents time.

We call X(t) the **state** of the process at time t.

If T is countably infinite, then we call **X** a **discrete time process**. We will mainly choose T to be the set of nonnegative integers.

Markov Chains

Definition

A discrete time process $\mathbf{X} = \{X_0, X_1, X_2, X_3, \ldots\}$ is called a Markov chain if and only if the state at time t merely depends on the state at time t - 1. More precisely, the transition probabilities

$$\Pr[X_t = a_t \mid X_{t-1} = a_{t-1}, \dots, X_0 = a_0] = \Pr[X_t = a_t \mid X_{t-1} = a_{t-1}]$$

for all values a_0, a_1, \ldots, a_t and all $t \ge 1$.

In other words, Markov chains are "memoryless" discrete time processes. This means that the current state (at time t - 1) is sufficient to determine the probability of the next state (at time t). All knowledge of the past states is comprised in the current state.

Homogeneous Markov Chains

Definition

A Markov chain is called **homogeneous** if and only if the transition probabilities are independent of the time t, that is, there exist constants $P_{i,j}$ such that

$$P_{i,j} = \Pr[X_t = j \mid X_{t-1} = i]$$

holds for all times t.

Assumption

We will assume that Markov chains are homogeneous unless stated otherwise.

Definition

We say that a Markov chain has a **discrete state space** if and only if the set of values of the random variables is countably infinite

 $\{v_0, v_1, v_2, \ldots\}.$

For ease of presentation we will assume that the discrete state space is given by the set of nonnegative integers

 $\{0, 1, 2, \ldots\}.$

Definition

We say that a Markov chain is **finite** if and only if the set of values of the random variables is a finite set

$$\{v_0, v_1, v_2, \ldots, v_{n-1}\}.$$

For ease of presentation we will assume that finite Markov chains have values in

$$\{0, 1, 2, \ldots, n-1\}.$$

The transition probabilities

$$P_{i,j} = \Pr[X_t = j \mid X_{t-1} = i].$$

determine the Markov chain. The transition matrix

$$P = (P_{i,j}) = \begin{pmatrix} P_{0,0} & P_{0,1} & \cdots & P_{0,j} & \cdots \\ P_{1,0} & P_{1,1} & \cdots & P_{1,j} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots \\ P_{i,0} & P_{i,1} & \cdots & P_{i,j} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots \end{pmatrix}$$

comprises all transition probabilities.

For any $m \ge 0$, we define the *m*-step transition probability

$$P_{i,j}^m = \Pr[X_{t+m} = j \mid X_t = i].$$

This is the probability that the chain moves from state i to state j in exactly m steps.

If $P = (P_{i,j})$ denotes the transition matrix, then the *m*-step transition matrix is given by

$$(P_{i,j}^m)=P^m.$$



A Markov chain with state space V and transition matrix P can be **represented** by a labeled directed graph G = (V, E), where edges are given by transitions with nonzero probability

$$E = \{(u, v) \mid P_{u,v} > 0\}.$$

The edge (u, v) is labeled by the probability $P_{u,v}$.

Self-loops are allowed in these directed graphs, since we might have $P_{u,u} > 0$.

Example of a Graphical Representation





Irreducible Markov Chains

We say that a state *j* is **accessible** from state *i* if and only if there exists some integer $n \ge 0$ such that

$$\mathsf{P}^n_{i,j} > 0.$$

If two states *i* and *j* are accessible from each other, then we say that they **communicate** and we write $i \leftrightarrow j$.

In the graph-representation of the chain, we have $i \leftrightarrow j$ if and only if there are directed paths from *i* to *j* and from *j* to *i*.

Proposition

The communication relation is an equivalence relation.

By definition, the communication relation is reflexive and symmetric. Transitivity follows by composing paths.

Definition

A Markov chain is called **irreducible** if and only if all states belong to one communication class. A Markov chain is called **reducible** if and only if there are two or more communication classes.

Proposition

A finite Markov chain is irreducible if and only if its graph representation is a strongly connected graph.



Question

Is this Markov chain irreducible?

$$P = \begin{pmatrix} 0 & \frac{1}{4} & 0 & \frac{3}{4} \\ \frac{1}{2} & 0 & \frac{1}{3} & \frac{1}{6} \\ 0 & 0 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$



Question

Is this Markov chain irreducible?

Answer

No, since no other state can be reached from 2.





Question

Is this Markov chain irreducible?





Question

Is this Markov chain irreducible?

Answer

Periodic and Aperiodic Markov Chains

Period

Definition

The **period** d(k) of a state k of a homogeneous Markov chain with transition matrix P is given by

$$d(k) = \gcd\{m \ge 1 \colon P_{k,k}^m > 0\}.$$

if d(k) = 1, then we call the state k aperiodic.

A Markov chain is **aperiodic** if and only if all its states are aperiodic.





Question





Question

What is the period of each state?

d(0) =





Question

What is the period of each state?

 $d(0) = 1, \ d(1) =$





Question

$$d(0) = 1, \ d(1) = 1, \ d(2) =$$





Question

$$d(0) = 1, d(1) = 1, d(2) = 1, d(3) =$$





Question

$$d(0) = 1$$
, $d(1) = 1$, $d(2) = 1$, $d(3) = 1$, so the chain is aperiodic.

Aperiodicity can lead to the following useful result.

Proposition

Suppose that we have an aperiodic Markov chain with finite state space and transition matrix P. Then there exists a positive integer N such that

 $(P^m)_{i,i} > 0$

for all states i and all $m \ge N$.

Before we prove this result, let us explore the claim in an exercise.



Question

What is the smallest number of steps N_i such that $P_{i,i}^m > 0$ for all $m \ge N$ for $i \in \{0, 1, 2, 3\}$?



Question

What is the smallest number of steps N_i such that $P_{i,i}^m > 0$ for all $m \ge N$ for $i \in \{0, 1, 2, 3\}$?

Answer $N_0 =$



Question

What is the smallest number of steps N_i such that $P_{i,i}^m > 0$ for all $m \ge N$ for $i \in \{0, 1, 2, 3\}$?

Answer

$$N_0 = 4$$
, $N_1 =$



Question

What is the smallest number of steps N_i such that $P_{i,i}^m > 0$ for all $m \ge N$ for $i \in \{0, 1, 2, 3\}$?

Answer

$$N_0 = 4$$
, $N_1 = 4$, $N_2 = 4$, $N_3 = 4$.

Now back to the general statement.

Proposition

Suppose that we have an aperiodic Markov chain with finite state space and transition matrix P. Then there exists a positive integer N such that

 $(P^m)_{i,i} > 0$

for all states i and all $m \ge N$.

Let us now prove this claim.

Proof.

We will use the following fact from number theory.

Lemma

If a subset A of the set of nonnegative integers is

- closed under addition, $A + A \subseteq A$, and
- satisfies $gcd\{a \mid a \in A\} = 1$,

then it contains all but finitely many nonnegative integers, so there exists a positive integer n such that $\{n, n + 1, n + 2, ...\} \subseteq A$.

Proof of the Lemma.

Suppose that $A = \{a_1, a_2, a_3, \ldots\}$. Since gcd A = 1, there must exist some postiive integer k such that

 $gcd(a_1, a_2, \ldots, a_k) = 1.$

Thus, there exist integers n_1, n_2, \ldots, n_k such that

 $n_1a_1+n_2a_2+\cdots+n_ka_k=1.$

We can split this sum into a positive part P and a negative part N such that

$$P-N=1.$$

As sums of elements in A, both P and N are contained in A.

Proof of the Lemma (Continued)

Suppose that *n* is a positive integer such that $n \ge N(N-1)$. We can express *n* in the form

$$n = aN + r$$

for some integer *a* and a nonnegative integer *r* in the range $0 \le r \le N - 1$.

We must have $a \ge N - 1$. Indeed, if a were less than N - 1, then we would have n = aN + r < N(N - 1), contradicting our choice of n.

We can express n in the form

$$n = aN + r = aN + r(P - N) = (a - r)N + rP.$$

Since $a \ge N - 1 \ge r$, the factor (a - r) is nonnegative. As N and P are in A, we must have $n = (a - r)N + rP \in A$.

We can conclude that all sufficiently large integers n are contained in A.
Proof of the Proposition.

For each state *i*, consider the set A_i of possible return times

$$A_i = \{m \ge 1 \mid P_{i,i}^m > 0\}.$$

Since the Markov chain in aperiodic, the state *i* is aperiodic, so $gcd A_i = 1$. If *m*, *m'* are elements of A_i , then

$$\Pr[X_m = i \mid X_0 = i] > 0$$
 and $\Pr[X_{m+m'} = i \mid X_m = i] > 0.$

Therefore,

$$\Pr[X_{m+m'} = i \mid X_0 = i] \ge \Pr[X_{m+m'} = i \mid X_m = i] \Pr[X_m = i \mid X_0 = i] > 0.$$

So m + m' is an element of A_i . Therefore, $A_i + A_i \subseteq A_i$.

By the lemma, A_i contains all but finitely many nonnegative integers. Therefore, A contains all but finitely many nonnegative integers. \Box

Let X be an irreducible and aperiodic Markov chain with finite state space and transition matrix P. Then there exists an $M < \infty$ such that $(P^m)_{i,j} > 0$ for all states i and j and all $m \ge M$.

In other words, in an irreducible, aperiodic, and finite Markov chain, one can reach each state from each other state in an arbitrary number of steps with a finite number of exceptions.

Proof.

Since the Markov chain is aperiodic, there exist a positive integer N such that $(P^n)_{i,i} > 0$ for all states *i* and all $n \ge N$.

Since *P* is irreducible, there exist a positive integer $n_{i,j}$ such that $P_{i,j}^{n_{i,j}} > 0$. After $m \ge N + n_{i,j}$ steps, we have

$$\underbrace{\Pr[X_m = j \mid X_0 = i]}_{P_{i,j}^m > 0} \ge \underbrace{\Pr[X_m = j \mid X_{m-n_{i,j}} = i]}_{=P_{i,j}^{n_{i,j}} > 0} \underbrace{\Pr[X_{m-n_{i,j}} = i \mid X_0 = i]}_{=P_{i,i}^{m-n_{i,j}} > 0}.$$

In other words, we have $P_{i,j}^m > 0$, as claimed.

Stationary Distributions

Definition

Suppose that X is a finite Markov chain with transition matrix P. A row vector $v = (p_0, p_1, \dots, p_{n-1})$ is called a **stationary distribution** for P if and only if

• the
$$p_k$$
 are nonnegative real numbers such that $\sum_{k=0}^{n-1} p_k = 1$.

$$vP = v.$$

Example

Every probability distribution on the states is a stationary probability distribution when P is the identity matrix.

Example

If
$$P = \begin{pmatrix} 1/2 & 1/2 \\ 1/10 & 9/10 \end{pmatrix}$$
, then $v = (1/6, 5/6)$ satisfies $vP = v$.

Any aperiodic and irreducible finite Markov chain has precisely one stationary distribution.

Total Variation Distance

Definition

If $p = (p_0, p_1, \dots, p_{n-1})$ and $q = (q_0, q_1, \dots, q_{n-1})$ are probability distributions on a finite state space, then

$$d_{TV}(p,q) = rac{1}{2}\sum_{k=0}^{n-1} |p_k - q_k|$$

is called the **total variation distance** between p and q.

In general, $0 \leq d_{TV}(p,q) \leq 1$. If p = q, then $d_{TV}(p,q) = 0$.

Definition

If $p^{(m)} = (p_0^{(m)}, p_1^{(m)}, \dots, p_{n-1}^{(m)})$ is a probability distribution for each $m \ge 1$ and $p = (p_0, p_1, \dots, p_{n-1})$ is a probability distribution, then we say that $p^{(m)}$ converges to p in total variation if and only if $\lim_{m \to \infty} d_{TV}(p^{(m)}, p) = 0.$

Let X be a finite irreducible aperiodic Markov chain with transition matrix P. If $p^{(0)}$ is some initial probability distribution on the states and p is a stationary distribution, then $p^{(m)} = p^{(0)}P^m$ converges in total variation to the stationary distribution,

$$\lim_{m\to\infty} d_{TV}(p^{(m)},p)=0.$$

Reversible Markov Chains

Definition

Suppose that **X** is a Markov chain with finite state space and transition matrix *P*. A probability distribution π on *S* is called **reversible** for the chain if and only if

$$\pi_i P_{i,j} = \pi_j P_{j,i}$$

holds for all states i and j in S.

A Markov chain is called **reversible** if and only if there exists a reversible distribution for it.

Suppose that **X** is a Markov chain with finite state space and transition matrix *P*. If π is a reversible distribution for the Markov chain, then it is also a stationary distribution for it.

Proof.

We need to show that $\pi P = \pi$. In other words, we need to show that

$$\pi_j = \sum_{k \in S} \pi_k P_{k,j}.$$

holds for all states *j*. This is straightforward, since

$$\pi_j = \pi_j \sum_{k \in S} P_{j,k} = \sum_{k \in S} \pi_j P_{j,k} = \sum_{k \in S} \pi_k P_{k,j},$$

where we used the reversibility condition $\pi_j P_{j,k} = \pi_k P_{k,j}$ in the last equality.

Random Walks

Definition

A random walk on an undirected graph G = (V, E) is given by the transitition matrix P with

$$P_{u,v} = \begin{cases} \frac{1}{d(u)} & \text{if } (u,v) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

For a random walk on a undirected graph with transition matrix P, we have

- P is irreducible if and only if G is connected,
- P is aperiodic if and only if G is not bipartite.

Proof.

If P is irreducible, then the graphical representation is a strongly connected directed graph, so the underlying undirected graph is connected. The converse is clear.

Proof. (Continued)

The Markov chain corresponding to a random walk on an undirected graph has either period 1 or 2. It has period 2 if and only if G is bipartite. In other words, P is aperiodic if and only if G is not bipartite.

A random walk on a graph with vertex set $V = \{v_1, v_2, ..., v_n\}$ is a Markov chain with reversible distribution

$$\pi = \left(\frac{d(v_1)}{d}, \frac{d(v_2)}{d}, \dots, \frac{d(v_n)}{d}\right),$$

where $d = \sum_{v \in V} d(v)$ is the total degree of the graph.

Proof.

Suppose that u and v are adjacent vertices. Then

$$\pi_u P_{u,v} = \frac{d(u)}{d} \frac{1}{d(u)} = \frac{1}{d} = \frac{d(v)}{d} \frac{1}{d(v)} = \pi_v P_{v,u}.$$

If u and v are non-adjacent vertices, then

$$\pi_u P_{u,v} = \mathbf{0} = \pi_v P_{v,u},$$

since $P_{u,v} = 0 = P_{v,u}$.

Example

$$|V| = 8$$
 and $|E| = 12$



Markov Chain Monte Carlo Algorithms

The Idea

Given a probability distribution π on a set S, we want to be able to sample from this probability distribution. In MCMC, we define a Markov chain that has π as a stationary distribution. We run the chain for some iterations and then sample from it.

The Idea

Given a probability distribution π on a set S, we want to be able to sample from this probability distribution. In MCMC, we define a Markov chain that has π as a stationary distribution. We run the chain for some iterations and then sample from it.

Why?

Sometimes it is easier to construct the Markov chain than the probability distribution π .

Hardcore Model

Definition

Let G = (V, E) be a graph. The hardcore model of G randomly assigns either 0 or 1 to each vertex such that no neighboring vertices both have the value 1.

Assignment of the values 0 or 1 to the vertices are called **configurations**. So a configuration is a map in $\{0, 1\}^V$.

A configuration is called **feasible** if and only if no adjacent vertices have the value 1.

In the hardcore model, the feasible configurations are chosen uniformly at random.

Question

For a given graph G, how can you directly choose a feasible configuration uniformly at random?

Question

For a given graph G, how can you directly choose a feasible configuration uniformly at random?

An equivalent question is:

Question

For a given graph G, how can you directly choose independent sets of G uniformly at random?

Grid Graph Example

Observation

In an $n \times n$ grid graph, there are 2^{n^2} configurations.

Grid Graph Example

Observation

In an $n \times n$ grid graph, there are 2^{n^2} configurations.

Observation

There are at least $2^{n^2/2}$ feasible configurations in the grid graph.

Indeed, set every other node in the grid graph to 0. For example, if we label the vertices by $\{(x, y) \mid 0 \le x < n, 0 \le y < n\}$. Then set all vertices with $x + y \equiv 0 \pmod{2}$ to 0. The value of the remaining $n^2/2$ vertices can be chosen arbitrarily, giving at least $2^{n^2/2}$ feasible configurations.

Direct sampling from the feasible configurations seems difficult.

Hardcore Model Markov Chain

Given a graph G = (V, E) with a set \mathcal{F} of feasible configurations. We can define a Markov chain with state space \mathcal{F} and the following transitions

- Let X_n be the current feasible configuration. Pick a vertex $v \in V$ uniformly at random.
- For all vertices $w \in V \setminus \{v\}$, the value of the configuration will not change: $X_{n+1}(w) = X_n(w)$.
- Toss a fair coin. If the coin shows heads and all neighbors of v have the value 0, then X_{n+1}(v) = 1; otherwise X_{n+1}(v) = 0.

Hardcore Model

Proposition

The hardcore model Markov chain is irreducible.

Hardcore Model

Proposition

The hardcore model Markov chain is irreducible.

Proof.

Given an arbitrary feasible configuration with m ones, it is possible to reach the configuration with all zeros in m steps. Similarly, it is possible to go from the zero configuration to an arbitrary feasible configuration with positive probability in a finite number of steps.

Therefore, it is possible to go from an arbitrary feasible configuration to another in a finite number of steps with positive probability.

The hardcore model Markov chain is aperiodic.

The hardcore model Markov chain is aperiodic.

Proof.

For each state, there is a small but nonzero probability that the Markov chain stays in the same state. Thus, each state is aperiodic. Therefore, the Markov chain is aperiodic.

Let π denote the uniform distribution on the set of feasible configurations \mathcal{F} . Let P denote the transition matrix. Then

$$\pi_f P_{f,g} = \pi_g P_{g,f}$$

for all feasible configurations f and g.

Proof.

Since $\pi_f = \pi_g = 1/|\mathcal{F}|$, it suffices to show that $P_{f,g} = P_{g,f}$.

• This is trivial if f = g.

• If f and g differ in more than one vertex, then $P_{f,g} = 0 = P_{g,f}.$

If f and g differ only on the vertex v. If G has k vertices, then

$$P_{f,g}=\frac{1}{2}\cdot\frac{1}{k}=P_{g,f}.$$
Corollary

The stationary distribution of the hardcore model Markov chain is the uniform distribution on the set of feasible configurations.